A Mixed Integer Approach for Obtaining Unique Solutions in Source Inversion of Drinking Water Networks[1]

Carl D. Laird, Lorenz T. Biegler[2]
Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
Bart G. van Bloemen Waanders
Sandia National Laboratories[3], Albuquerque, NM, 87109, USA

1 Introduction

Presidential Decision Directive 63 identified water systems as one of the critical infrastructures to the United States. Following this directive and the passing of *The Public Health, Security, and Bioterrorism Preparedness and Response Act* there has been increased research effort in both assessing the vulnerability of drinking water systems and proposing protection measures. Drinking water networks are vulnerable to chemical and biological contamination. While physical security is being used to limit access to some potential contamination locations, due to the distributed nature of drinking water networks, many locations remain unprotected. One proposed method of protection is the installation of an early warning detection system. Sensors installed at various locations throughout the drinking water network could warn utilities companies in the event of a contamination.

On its own, an early warning detection system [7, 1], provides only a coarse measure of the time and location of the contamination event. In previous work [5, 4, 13] the authors introduced a large scale nonlinear programming approach that used real-time concentration information from an installed sensor grid to accurately determine the time and location of the contamination event. This approach introduced unknown, time dependent injection terms at every node in the network and formulated a quadratic program to solve for the time profiles of the injections. The objective function was a least squares minimization of the errors between the calculated and measured node concentrations at the sensor nodes with a regularization term to force a unique solution. The constraints in the optimization problem were the partial differential equations of the water quality model for the network. In Laird et al. [5] this problem was then discretized with a fully simultaneous approach, using an *origin tracking algorithm* to characterize the pipe time delays and remove the need to discretize along the length of the pipes. The resulting large scale nonlinear program was solved using a nonlinear interior point code, IPOPT [14]. This approach was effective at identifying a family of possible injection scenarios.

The unregularized formulation of the source inversion problem can have many non-unique solutions. The regularized formulation, on the other hand, has a unique solution, but this solution is essentially linear combination of possible injection scenarios. With this approach alone, it is difficult to determine if the observed contamination was caused from a single injection location or multiple locations. In this work, we propose a problem reduction technique and formulate a mixed integer quadratic program (MIQP) to identify unique injection scenarios. This formulation includes constraints that further limit the solution space and allows us to distinguish between single and multiple injection locations.

Section 2 gives a brief description of the formulation presented in [5], followed by a discussion of solution non-uniqueness and how this manifests in the regularized problem. In Section 3 we introduce the mixed integer formulation and show how the problem size can be reduced drastically

[1]Financial support for this work was provided by the National Science Foundation under ITR Grant ACI-021667.
[2]Corresponding author, email: lb01@andrew.cmu.edu
[3]Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

using active-set information from the original continuous problem. We show the effectiveness of this approach on a real municipal water network in Section 4. Here we test both single and multiple location injection scenarios. Finally, we present some conclusions and directions for future work.

2 Background

The purpose of the work in [5] was to present a formulation for solving the inverse problem of identifying the time and location of contaminant injections in real-time, using concentration information from a sparse sensor grid. This work assumed that contaminations occurred at network nodes and introduced unknown, time dependent contaminant injections at each node in the network. The optimization problem was then written as a weighted least squares minimization of the errors between the calculated and measured concentrations subject to the constraints of the water quality model. As with water quality simulation techniques [12, 15, 10], this approach assumed that the network flows are known inputs and modeled the water quality only. The solution to the least squares problem provides the complete time dependent profiles of the unknown mass injections at every node that give rise to the best weighted least squares match of the observed data.

In Laird et al. [5] an origin tracking algorithm was presented to characterize the pipe time delays and remove the need to discretize in space. Following this reduction, the remaining problem was then discretized in time alone, producing a nonlinear program of reasonable size. Using Θ to refer to the set of discretizations in time, and $\mathcal{P}$ and $\mathcal{N}$ to refer to the compete set of network pipes and nodes, respectively, the original continuous quadratic program (OCQP) for the inversion problem can be formulated as [5],

$$\min_{\bar{c},c,m} \quad f = \frac{1}{2}\left[c - c^{\star}\right]^T W \left[c - c^{\star}\right] + \frac{\rho}{2} m^T m \tag{1}$$

$$\text{s.t. } \bar{c} - Pc = 0, \tag{2}$$

$$\bar{N}\bar{c} + Nc + Mm = 0, \tag{3}$$

$$m \geq 0, \tag{4}$$

where $\bar{c}=\left[\cdots \bar{c}^I_{i,j}, \bar{c}^O_{i,j} \cdots\right], \forall i \in \mathcal{P}, j \in \Theta$ is a vector of pipe concentrations for the inlet ($\mathcal{I}$) and outlet ($\mathcal{O}$) of every pipe, discretized in time, $c=\left[\cdots c_{i,j} \cdots\right], \forall i \in \mathcal{N}, j \in \Theta$ is the vector of calculated concentrations for every node at every time discretization, and $m=\left[\cdots m_{i,j} \cdots\right], \forall i \in \mathcal{N}, j \in \Theta$ is the vector of unknown contaminant mass injections for every node at every time discretization. The matrix P is defined by the origin tracking algorithm, and the matrices $\bar{N}$, N, and M are the Jacobians of the discretized mass balance constraints (for junctions and storage tanks) with respect to the pipe concentrations, node concentrations, and unknown injections, respectively. The diagonal matrix W is a weighting matrix for the least squares errors, with nonzero elements only for sensor nodes at sample times. The integral of the absolute value of flow through the corresponding node is used for the nonzero elements. This provides a flow based weighting and shifts the least squares error from a concentration basis to a mass basis.

The second term in the objective is a regularization term that forces a unique solution to the problem. First, consider the unregularized case. With ρ=0 problem (1-4) is a convex (not necessarily strictly convex) quadratic program; therefore, the QP solution is a global minimum for the objective value, but the minimizer it is not necessarily unique. Consider the grid network shown in Figure 1 with sensors installed at every second node (indicated by the grey shading) and a single reservoir water source at node 26. Network flows are constant in the directions indicated by the arrows with demands at the boundary nodes only. These demands were selected to introduce symmetry

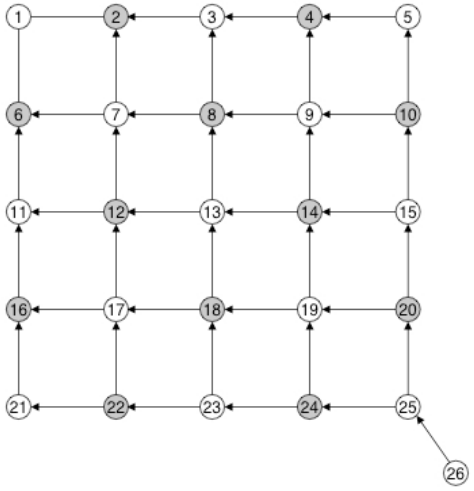

Figure 1: Grid Network Example. A small symmetric grid network with sensors installed at every second node, indicated by the shading.

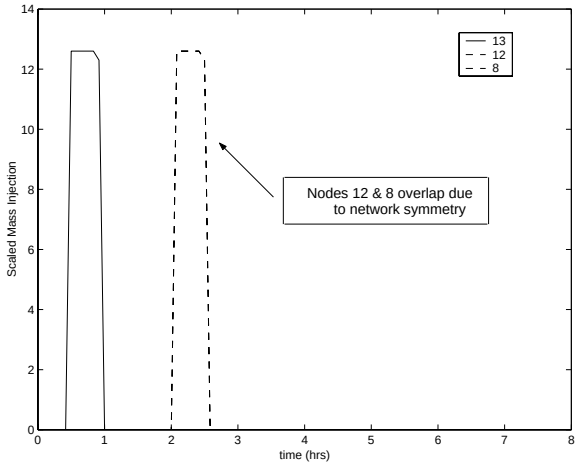

Figure 2: Grid Network Example Solution. Solution of problem (1-4) on the grid network with an injection from node 13 at $t = 0.5$ hours.

about an axis through nodes 1 and 26 and time delays that range 0.5 to 5 hours. Selecting node 13 as the injection location, we introduce contaminant from time, $t = 0.5$ to $t = 1$ hours. With the flows used in this network, the contaminant will flow from node 13 to nodes 12 and 8 in about an hour and a half, where the sensors will then detect the contaminant. If we consider only the observed concentration measurements at nodes 8 and 12 (and zero concentration measurements from the other sensors), we see two possible injection scenarios capable of producing the observed concentrations. The contaminant could have been injected at node 13 (the actual injection) or at both nodes 8 and 12 simultaneously. As well, any linear combination of these two scenarios is also a possible solution to the unregularized problem.

There are an infinite number of solutions to the unregularized problem. With a small positive value for ρ, however, the QP obtains a unique *regularized* solution. Simulating the injection from node 13 with EPANET [11], Laird et al. [5] performed the origin tracking algorithm and, using the EPANET concentrations as sensor measurements, formulated problem (1-4) with ρ=$1 \cdot 10^{-4}$, 5 minute timesteps, and a time horizon of 4 hours (48 timesteps). The optimization problem was written in AMPL [3] format by a software tool that reads the EPANET input and output files. This AMPL problem was then solved using IPOPT [14] and the solution is shown in Figure 2. This figure shows the time profiles for the non-zero mass injections only with time in hours along the abscissa and the mass flowrate of the injections along the ordinate. As expected, we see that nodes 8, 12, and 13 all contribute injection profiles to the solution. The effect of the regularization term is apparent. The solution given is a linear combination of the two solutions discussed above. With the simplicity of the grid network topology and flow patterns, it is straightforward for us to examine Figure 2 and see the two solutions of interest, namely an injection at node 13 at $t = 0.5$ hours, or a simultaneous injection at nodes 8 and 12 at $t = 2$ hours. Unfortunately, for a more complex network, with time varying flow patterns, interpreting these solution profiles is non-trivial.

The existence of non-unique solutions is unavoidable, as they are a direct function of the incomplete sensor configuration. The regularized OCQP gives only one regularized solution. The unique scenarios that give rise to this linear combination need to be inferred. Knowing only that the contaminant could have been injection at node 8, 12, or 13 (or any combination) does not allow us to

prioritize the locations or determine their individual likelihood. Given the knowledge that a single injection from node 13 could produce the observed concentrations, we would likely investigate this location first, as opposed to nodes 8 and 12, where simultaneous injections are required. Furthermore, while it is important to identify when a contamination event could have occurred from a single location, it is even more important to identify when this is not possible. We are particularly interested in immediately identifying when contamination events involve multiple injection locations.

In the next section, we introduce an approach that not only indicates when multiple injections are likely, it also provides the time and location of each injection. We propose an algorithm that uses a mixed integer quadratic programming (MIQP) formulation to search the space of solutions provided by the OCQP given by (1)-(4). The mixed integer formulation allows us to include discrete constraints on the problem, such as limiting the number of injection locations and provides more information than the OCQP solution alone.

3 Mixed Integer Formulation

Consider the mixed integer reformulation of (1)-(4),

$$\min_{\bar{c},c,m} \quad f = \frac{1}{2}\left[c - c^\star\right]^T W \left[c - c^\star\right] \tag{5}$$

$$\text{s.t.} \qquad \bar{c} - Pc = 0, \tag{6}$$

$$\bar{N}\bar{c} + Nc + Mm = 0, \tag{7}$$

$$Ly_i \leq m_{i,j} \leq Uy_i, \qquad \forall i \in \mathcal{N}, j \in \Theta, \tag{8}$$

$$\sum_{i \in \mathcal{N}} y_i = n. \tag{9}$$

Here, $y_i \in \{0, 1\}$ is a binary variable, n is a parameter representing the number of injection locations. The parameters L and U correspond to the nonzero detection limit and an upper bound on the injection variables, respectively. Problem (5)-(9) has the same number of continuous variables as the OCQP and the number of binary variables equals the number of nodes in the network. However, since our goal is to search the space of solutions arising from the OCQP, we do not need to formulate the MIQP with the full space of variables.

Consider the grid network problem presented in Section 2. With 26 nodes and 48 timesteps, this problem has 1248 continuous mass injection variables (one for each node at each timestep). If we look at the solution of the OCQP given in Figure 2, we immediately notice that most of the bound constraints (4) are active, meaning that the majority of the mass injection variables are at their lower bound of zero. In this example, there are only 18 nonzero mass injection variables (6 timesteps for each of the 3 nodes). Furthermore, there are only 3 nodes represented in the solution, 8, 12, and 13. To search within the solution space of the OCQP, we only need to consider nonzero mass injection variables and only the reduced set of node locations.

Therefore, if we first solve the OCQP and identify the injection variables that are zero (or below some detection limit), we can remove these variables and formulate the mixed integer problem in the space of the nonzero variables only. Let m_r refer to the set of mass injection variables that were inactive[4] at the solution of the OCQP. The set $\mathcal{N}_r$ contains all the nodes represented by this selection of mass injection variables. The set $(\Theta_r)_i$ contains the timesteps of the selected injection

[4]With an interior point method, the active variables will not be pushed completely to their bounds and we must select the active set using a small tolerance.

variables for node i. With this notation, $m_r = \left[\cdots (m_r)_{i,j} \cdots\right], \forall i \in \mathcal{N}_r, j \in (\Theta_r)_i$. Defining M_r as the columns of M in the space of m_r, we let $A_r{=}(\bar{N} \cdot P + N)^{-1} M_r$, we further reduce the problem by removing the pipe and node concentrations. Defining $m_r^\star$ as the solution of the original problem in the space of the selected variables only, we introduce a new variable, $\Delta m_r{=}m_r - m_r^\star$. Since the value of the regularization parameter ρ is small, we can neglect the regularization term and, using the reductions described above, rewrite the objective function (1) as,

$$\begin{aligned}
&\frac{1}{2}\left[A_r(m_r^\star + \Delta m_r) - c^\star\right]^T W \left[A_r(m_r^\star + \Delta m_r) - c^\star\right] \\
&= \frac{1}{2}\left[A_r m_r^\star - c^\star\right]^T W \left[A_r m_r^\star - c^\star\right] \\
&\qquad\qquad + \left[A_r m_r^\star - c^\star\right]^T W \left[\Delta m_r\right] + \frac{1}{2}\left[A_r \Delta m_r\right]^T W \left[A_r \Delta m_r\right]. \quad (10)
\end{aligned}$$

The term $\frac{1}{2}\left[A_r m_r^\star - c^\star\right]^T W \left[A_r m_r^\star - c^\star\right]$ is constant and, from the optimality conditions for the original problem, we know that $\left[A_r m_r^\star - c^\star\right]^T W{=}0$ (the multipliers for the inactive bound constraints are all zero). Therefore, both of these terms can be removed from the objective and we write the reduced form of problem (5)-(9) as,

$$\min_{m_r} \quad \hat{f} = \frac{1}{2}\left[\Delta m_r\right]^T Q \left[\Delta m_r\right] \quad (11)$$

$$\text{s.t.} \qquad L y_i \le \left(m_r^\star + \Delta m_r\right)_{i,j} \le U y_i, \qquad \forall i \in \mathcal{N}_r, j \in (\Theta_r)_i, \quad (12)$$

$$\sum_{i \in \mathcal{N}} y_i = n, \quad (13)$$

where, $Q{=}A_r^T W A_r$. The matrix A_r can be calculated by perturbing each element of m_r and simulating to find the response on c. Formulation (11-13) is a Mixed Integer Quadratic Program (MIQP) and forms the base problem for our search of solutions.

In addition to constraint (13) which limits the space of solutions to those with n injection locations, we can also add integer cuts [2] eliminating unlikely scenarios or previously visited solutions. To cut a particular combination of injections from the solution space, define $\mathcal{Y}_1$ to be the set of y_i variables that are one and $\mathcal{Y}_0$ to be the set of binary variables that are zero. We can then add the cut,

$$\sum_{i \in \mathcal{Y}_1} y_i - \sum_{i \in \mathcal{Y}_0} y_i = |\mathcal{Y}_1| - 1, \quad (14)$$

to remove that particular combination from the set of possible solutions. Here, $|\mathcal{Y}_1|$ refers to the cardinality of $\mathcal{Y}_1$ (the number of elements in the set).

To show the effectiveness of this MIQP formulation, we again consider the grid network example from Section 2 and formulate problem (11-13) using the solution of the OCQP, shown in Figure 2. Since the number of injection locations in the solution is only 3, we can completely enumerate all the possible combinations for the binary variables.

Table 1 shows the optimal objective value for each combination of binary variables, sorted first by objective value and then by $\sum_{i \in \mathcal{N}_r} y_i$. Notice that the top entry in the table tells us immediately that an injection at node 13 alone is able to reproduce the observed data precisely. Considering the possibility of two injection locations, injections at node 8 and 12 together are also able to reproduce

Table 1: Enumerated Solutions to Grid Network Example.

Objective Value	n	y_8	y_{12}	y_{13}
0.000	1	0	0	1
0.000	2	1	1	0
0.000	3	1	1	1
1.439	2	0	1	1
1.439	2	1	0	1
169.9	1	1	0	0
169.9	1	0	1	0
339.7	0	0	0	0

the data precisely. These numbers confirm what we already know from the network topology and flow conditions. Of course, the solution given by the OCQP, with all three injection locations, is also capable of matching the data precisely. It appears as though we get a reasonable match for a solution with injections at both nodes 12 and 13. In this case, the mass injection profile (not shown) for node 12 is at its lower bound, L for all timesteps considered. The solution shown here is matching the data well with the injection at node 13, and then adding as little as possible from node 12. Increasing the value of L will force an increase in the objective value for this case. The same is true of the case with node 8 and 13. Nevertheless, we need not be concerned since better objective matches have been seen with node 13 alone. Table 1 also shows that a single injection at node 8 or node 12 is not able to match the observed data.

For larger problems, we do not enumerate all the possible combinations. Instead, we set a particular value for the number of injection locations, n, and evaluate the MIQP repeatedly, searching for good solutions. Starting with n=1 we find all solutions that provide a reasonable match, excluding each previous solution as we proceed. We then increase n and continue. Our recommended approach for solving the source inversion problem and identifying the significance of solutions is as follows.

Algorithm 1

1. Solve the OCQP:
   Given sensor and flow data, formulate and solve the OCQP, as described in Laird et al. [4, 5]. Identify the set of selected nodes and timesteps for $\mathcal{N}_r$ and $(\Theta_r)_i, \forall i \in \mathcal{N}_r$.

2. Calculate Q:
   Each column in A_r is the gradient of the vector c with respect to a particular entry in m_r. Set all the entries of m_r to zero except for one, and solve the forward problem. The values of c from the simulation form a column of A_r. Performing the perturbation for each $i \in \mathcal{N}_r, j \in (\Theta_r)_i$ gives us the entire matrix A_r. Calculate Q=$A_r^T W A_r$.

3. Formulate the base MIQP:
   Select reasonable upper and lower bounds U, and L, for the injection variables and formulate the mixed integer quadratic program (11-13). This problem forms the base MIQP.

4. Find reasonable upper bound on the objective, $\hat{f}$:
   Force all injections to zero by setting n=0. Solve the base MIQP to find $\hat{f}$ for the case where injection variables are all zero. Since this represents a reasonable upper bound on the objective, store the objective value as $\hat{f}_0$.

5. for each n=1..$|\mathcal{N}_r|$:

   5.1. Initialize the problem for the current value of n:
   Set $\hat{f}$=0 and formulate the base MIQP for n injection locations.

   5.2. while ($\hat{f}/\hat{f}_0 < \epsilon$[5]):

      5.2.1. Solve the MIQP:
      Solve problem (11)-(13) with any additional integer cuts and record the values of the binary variables and the objective. Store the value of the objective in $\hat{f}$.

      5.2.2. Exclude previous solution:
      Add the integer cut (14) for the values of the binary variables from the last solution. Return to Step 5.2 to continue finding new solutions for the current value of n.

Upon completion of this algorithm all the possible solutions with objective values less than ten percent of $\hat{f}_0$ will be recorded. One can then examine these solutions to determine which injections correspond to likely scenarios.

4 Results

In this section, we demonstrate the effectiveness of Algorithm 1 and the MIQP formulation using a real municipal water network model. We simulate two injection scenarios, one with a single injection location and one with two injection locations. We then perform Algorithm 1 on both these scenarios, demonstrating that this approach is capable of distinguishing between the single injection scenario and the multiple injection scenario.

We test the approach on a real municipal water network with approximately 500 nodes and 600 links. A nodally perturbed version of the network is shown in Figure 3. We select 50 sensor nodes based on a weighted random selection. In this selection, we first assign a weight to each node, using the total volume of water that has flowed through each node over a simulated 16 hour period. To ensure we do not skew the selection drastically, any node with a weight that is less than five percent of the maximum weight is set to this lower bound. We then randomly sample this distribution 50 times to select the sensor nodes.

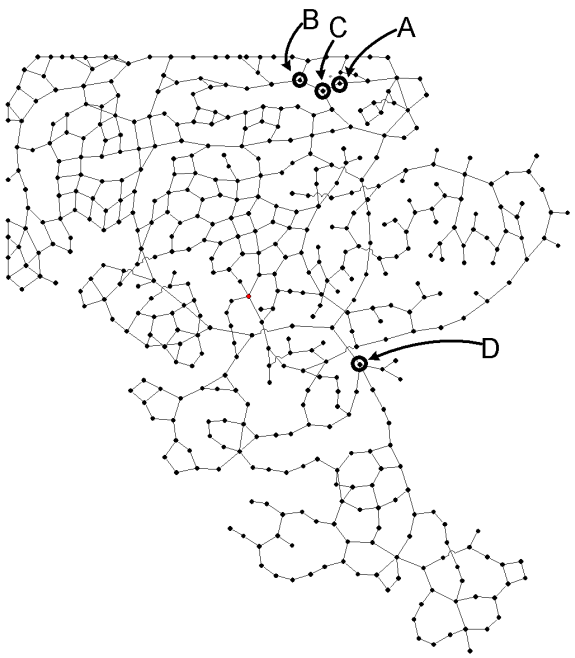

Figure 3: Real Municipal Water Network: The network model showing the four nodes of interest, A, B, C, and D. The physical locations of all the network nodes have been shifted, but the connectivity remains true.

In Test 1, we simulate an hour long contamination from node A at t=1 hour. In Test 2, we keep this contamination event and add another hour long injection from node D, also at time t=1 hour. Performing Algorithm 1, we first formulate and solve the OCQP. The solution of OCQP for both of these tests is not shown since they contain such a large number of injection nodes. For Test 1,

[5]The parameter ϵ represents the limit on $\hat{f}/\hat{f}_0$ and is set to 0.1 in this study.

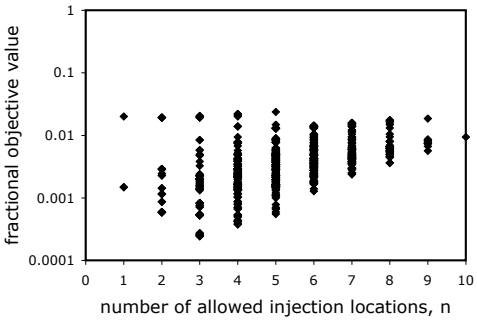

Figure 4: Recorded Solutions for Test 1. Results of Algorithm 1 for the contamination scenario with a single injection location.

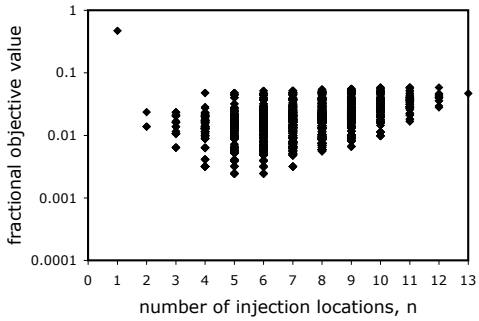

Figure 5: Recorded Solutions for Test 2. Results of Algorithm 1 for the contamination scenario with two injection locations.

the solution of the OCQP indicates 10 nodes and 73 injection variables to consider (3 nodes with 13 timesteps, 6 nodes with 5 timesteps, and one node with 4 timesteps). In Test 2, the scenario with two injection locations, there are 13 nodes and 79 injection variables to consider (one node with 13 timesteps, 3 nodes with 12 timesteps, 2 nodes with 5 timesteps, 4 nodes with 4 timesteps, and 3 nodes with 2 timesteps).

Using the solutions from the OCQP we formulate the base MIQP in AMPL [3] format. We then continue the algorithm, using the NEOS Server for Optimization (www-neos.mcs.anl.gov) where each MIQP was solved with the mixed integer, nonlinearly constrained algorithm, MINLP branch and bound solver by Fletcher and Leyffer [6].

Figures 4 and 5 show all the recorded solutions from Algorithm 1. The horizontal axis shows n, the number of allowed injection locations, and the vertical axis shows the fractional objective value for each solution on a log scale. Many different solutions may have the same objective values and points may overlap in these figures. With the approach outlined in Algorithm 1, we have a column of solutions for each value of n. For Test 1 there are single injection solutions that provide a reasonable match. For Test 2, the two injection contamination scenario, there are no single injection solutions that match the observed data. This immediately tells us that we have a situation with multiple injections and demonstrates the ability of this approach to differentiate single location contaminations from multi-location contaminations.

While Figures 4 and 5 indicate the number of likely injection locations, they do not show us the nodes involved in the injections. This information was recorded in the algorithm, however, and Tables 2 and 3 summarize the solutions near the lower left corner of Figures 4 and 5. The results in Table 2 show three single injection solutions that match the observed data (the fractional objective values are all under one percent). These three are neighboring nodes in the network and, with no sensor between them, are indistinguishable with the observed data. These results indicate that a single injection location is possible at any one of these three locations. Even more encouraging are the results for Test 2, shown in Table 3. In this test, we simulated two injections, one from Node A and one from node D. The results show that there are no single injection scenarios that could have produced the observed data (the best fractional objective for a single injection solution is almost fifty percent). On the contrary, there are three solutions with two injection locations that match the data. Table 3 shows that all three solutions with two injection locations include node D. We also know from Test 1 that nodes A, B, and C are indistinguishable with the selected sensor layout. It is not surprising then, that the three solutions that match the data are node A with D, node

Table 2: Detailed Results for Test 1

n	Frac. Obj.	Comment
1	0.001	y_A=1, others 0
1	0.001	y_C=1, others 0
1	0.020	y_B=1, others 0
2	0.001	y_A=1[a]
⋮	⋮	⋮

[a]One of the other nodes, not shown in Figure 3 has a value of 1, all others 0.

Table 3: Detailed Results for Test 2

n	Frac. Obj.	Comment
1	0.473	y_D=1, others 0
2	0.014	y_C=y_D=1, others 0
2	0.014	y_A=y_D=1, others 0
2	0.024	y_B=y_D=1, others 0
2	0.465	y_D=1[a]

[a]One of the other nodes, not shown in Figure 3 has a value of 1, all others 0.

B with D, and node C with D. Considering only the results from the solution of the OCQP, we could not determine if the contamination was the result of a single injection or multiple injections. With the mixed integer analysis on Test 2, however, we can immediately see that there are multiple injection sites to consider, and, with surety, we know which sites we should immediately investigate.

5 Conclusions and Future Work

These results demonstrate that the MIQP formulation and the algorithm presented are extremely effective at identifying unique solutions from the family of solutions provided by the original problem (OCQP). The approach is able to identify unique solutions and find good solutions with a minimum number of injection locations. The greatest strength of this approach, however, is its ability to distinguish multiple injections. The results clearly show, for the scenarios considered, that the two injection scenario was distinguishable from single injection scenario and the injection locations were identified as accurately as possible with the sparse sensor layout.

Furthermore, the general MIQP takes only minutes to formulate (dominated by the calculation of Q) with a computational complexity that is linear in the number of selected injection variables. Tables 2 and 3 only show a portion of the number of problems that were solved. For completeness, we enumerated the entire set of possible binary variables and each MIQP solved in under a second. These favorable computation results make this approach reasonable for a real-time setting.

While these results are very encouraging, both in their effectiveness and their computational effort, they only represent two injection scenarios. It remains for us to analyze the behavior of the MIQP approach and Algorithm 1 on additional injection scenarios. It will also be interesting to see how well the approach can differentiate injections from more than two locations.

In all tests, the sensor data was simulated and the demands were characterized perfectly. In a real life situation, the sensor data will be noisy and the demands will only be loosely characterized. We need to develop a framework to perform the source inversion considering the uncertainty associated with these parameters. For this we propose to use a multi-scenario framework, where a single optimization problem is formulated with a large number of scenarios, each being a statistical sample from the expected demand distributions. The injection variables will be common across each of these scenarios and a solution of this problem will yield an expected value for the injection variables under the uncertainty of the demands. This approach has been previously described for the design of chemical plants under uncertain conditions in Rooney and Biegler [9]. While the optimization problem grows linearly with the number of scenarios considered, the structure of the linear system solved at each iteration of the optimization can be exploited to provide efficient solutions for a very large number of scenarios. Furthermore, great performance gains can be made by solving this system in parallel. We believe that this is the next critical step in solving the problem of effective

source inversion in drinking water networks.

6 Acknowledgments

The authors would like to acknowledge Dr. Feng Shang whose conversation renewed our interest in this problem. This discussion encouraged us to reconsider problem reduction techniques, allowing the development of the work within this paper.

References

[1] J. Berry, L. Fleischer, W. Hart, and C. Phillips. Sensor placement in municipal water networks. *Proceedings of the EWRI Conference, Philadelphia*, 2003.

[2] Lorenz T. Biegler, Ignacio E. Grossmann, and Arthur W. Westerberg. *Systematic Methods of Chemical Process Design*. Prentice Hall International Series in the Physical and Chemical Engineering Sciences. Prentice Hall PTR, Upper Saddle River, New Jersey 07458, 1997.

[3] R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming*. The Scientific Press (now an imprint of Boyd & Fraser Publishing Co.), Danvers, MA, USA, 1993. ISBN 0-89426-232-7.

[4] C. D. Laird, L. T. Biegler, and B. van Bloemen Waanders. Real-time, large scale optimization of water network systems using a subdomain approach. *submitted for publication, presented at Second CSRI Conference on PDE-Constrained Optimization, Santa Fe, NM*, 2004.

[5] C.D. Laird, L.T. Biegler, B.G. van Bloemen Waanders, and R.A. Bartlett. Contamination source determination for water networks. *A.S.C.E. Journal of Water Resources Planning and Management*, March/April 2005.

[6] S. Leyffer. Integrating sqp and branch-and-bound for mixed integer nonlinear programming. *Computational Optimization and Applications*, 18:295–305, 2001.

[7] A. Ostfeld and E. Salomons. An early warning detection system (ewds) for drinking water distribution systems security. *Proceedings of the EWRI Conference, Philadelphia*, 2003.

[8] President William J. Clinton Presidential Decision Directive. *Critical Infrastructure Protection: Sector Coordinators, Presidential Decision Directive 63*. Federal Register 63, no. 150 (5 August 1998): 41804.

[9] W. C. Rooney and L. T. Biegler. Optimal process design with model parameter uncertainty and process variability. *AIChE Journal*, 49(2):438, 2003.

[10] L.A. Rossman and P.F. Boulos. Numerical methods for modeling water quality in distribution systems: A comparison. *Journal of Water Resources PLanning and Management*, 122(2):137–146, March/April 1996.

[11] L.W. Rossman. Epanet 2. Technical report, EPA, 2000.

[12] F. Shang, J. Uber, and M. Polycarpou. A particle backtracking algorithm for water distribution systems analysis. *A.S.C.E. Journal of Environmental Eng*, 128(5):441–450, May 2002.

[13] B. G. van Bloemen Waanders, R. A. Bartlett, L. T. Biegler, and C. D. Laird. Nonlinear programming strategies for source detection of municipal water networks. *Presented at EWRI Conference, Philadelphia*, 2003.

[14] A. W achter and L. T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Accepted for publication in Mathematical Programming*, to appear March 2004.

[15] M.L. Zierolf, M.M. Polycarpou, and J.G. Uber. Development and auto-calibration of an input-output model of chlorine transport in drinking water distribution systems. In *I.E.E.E. Trans. on Control Systems Technology*, volume 6(4), pages 543–553, 1998.